Europ. J. Combinatorics (1991) **12**, 331–339

Finite Codes and Groupoid Words

JONATHAN D. H. SMITH

Correspondences between codes and groupoid words are established: between maximal suffix codes and groupoid words; between maximal commutatively prefix codes and cancellative entropic groupoid words; and between finite maximal codes and idempotent entropic groupoid words. Recognition criteria for the three types of groupoid word are given in terms of order relations involving binomial coefficients, divisibility conditions, partition functions and curves in simplices.

1. INTRODUCTION

The primary aim of this paper is to examine various correspondences between groupoid words and codes in a binary alphabet. Roughly speaking, the correspondences connect groupoid words with suffix (or prefix) codes (Section 2), cancellative entropic groupoid words with commutatively prefix codes (Section 3), and idempotent entropic groupoid words with finite maximal codes (Section 6). Codes, as bases for free semigroups, lie within the domain of the associative law. On the other hand, the groupoids of interest are non-associative. Thus the correspondences offer some curious connections between associative and non-associative phenomena.

From the algebraic side, the main concern is with 'recognition criteria'. Free groupoids of various types appear naturally as subsets of larger sets with algebraic structure. The 'recognition problem' is to determine when elements of these larger sets actually lie within the subsets representing the free groupoids. The standard recognition criteria given in the algebraic literature are often low-level syntactical criteria, while coding theory offers more elegant, higher-level semantical criteria. The correspondences discussed enable one to translate these more amenable criteria to the algebraic context. In Section 2, semiterms are recognized as representing groupoid words precisely when their monomials form a maximal suffix code. (Suffix codes appear following the parenthesis-minimizing algebraic convention of composing maps in the natural reading direction from left to right: the opposite convention would yield prefix codes.) Theorem 3.3 gives a divisibility criterion for recognizing which elements of free modules over bivariate polynomial rings represent cancellative entropic groupoid words in the free generators. This divisibility condition (3.5) emerges as equivalent to the rather complicated order conditions on binomial coefficients (3.4) known in the algebraic literature. Theorem 4.1 gives a comparable divisibility criterion in terms of polynomials in a single variable. Section 5 discusses recognition criteria for idempotent entropic groupoid words within modules over polynomial rings in a single variable. The criteria here come from abstract convexity theory rather than from coding theory. One criterion (Theorem 5.1) is syntactical, involving order relations with binomial coefficients, while the other (Theorem 5.2) is geometric, correlating idempotent entropic groupoid words with curves from one corner of a simplex to another.

From the coding theory point of view, the correspondences focus attention toward algebraic structure of binary codes that is often overlooked; for example, the free groupoid structure formed by finite maximal suffix codes (Proposition 2.2). Commutative equivalence of codes makes especial sense in terms of the map k of (6.4), while

0195-6698/91/040331 + 09 $02.00/0

measures of finite codes, particularly finite maximal codes, are conveniently viewed purely algebraically in terms of the map m of (6.5) or the composite km. Hovering in the background is the problem of determining whether every finite maximal code is commutatively prefix. Section 6 approaches this problem for binary codes (how closely remains to be seen) in terms of lifting the recognition criteria of Section 5 for idempotent entropic words back along the map m to the recognition criteria of Section 3 for cancellative entropic words. Finally, Section 4 gives a statistical mechanical interpretation to Karp's univariate structure function used to recognize finite languages commutatively equivalent to maximal prefix codes. Under the correspondence of Section 3, this interpretation carries over to cancellative entropic groupoid words. There is much current interest in physical interpretations of knot polynomials (e.g. [10]). Now certain knot polynomials are naturally realized as entropic right quasigroup words [8]. Entropic right quasigroups have two interconnected groupoid operations. Part of the motivation for the present study has been the desire to clarify aspects of entropic groupoids as a preparation for understanding those aspects of entropic right quasigroups that are relevant to knot theory. In particular, the physical interpretation given to entropic groupoid words in Section 4 offers clues for a physical interpretation of knot polynomials via entropic right quasigroup words.

2. Groupoids and Maximal Suffix Codes

A *groupoid* or *magma* $(G, \circ)$ is a set G equipped with a binary operation of *multiplication* denoted by infixed $\circ$ or juxtaposition (with juxtaposition binding more strongly). Given a set A, often referred to as an *alphabet* with elements called *letters*, the set A° of *groupoid words* in A is defined inductively as follows:

$$\left.\begin{array}{l}\text{(i) each letter is a groupoid word;}\\ \text{(ii) if } u \text{ and } v \text{ are groupoid words, then so is } u \circ v.\end{array}\right\} \tag{2.1}$$

By (ii), the operation $\circ$ makes A° a groupoid, the *free groupoid* on A. (This terminology is justified by Schröter's Theorem [7, 133].) Groupoid words in A may be represented by rooted binary trees with labelled edges and leaves. The edges are labelled by elements of the set $\{R, L\}$ and the leaves are labelled by elements of the set A. The representing tree or *parsing tree* T_w of a groupoid word w in A is defined inductively as follows:

$$\left.\begin{array}{l}\text{(i) each letter is represented by a single vertex labelled by the letter;}\\ \text{(ii) } T_{u \circ v} \text{ consists of } T_u,\ T_v, \text{ and a new root which is joined to the root of } T_u \text{ by an}\\ \text{edge labelled } R \text{ and to the root of } T_v \text{ by an edge labelled } L.\end{array}\right\} \tag{2.2}$$

Given a vertex in such a parsing tree, the unique directed geodesic from the vertex to the root is described by the concatenated sequence of edge labels along the geodesic. If the given vertex is the root itself, the corresponding sequence is taken to be 1, the empty concatenation. In this way the geodesics and their initial vertices are specified by elements of the free monoid $\{R, L\}^*$ over $\{R, L\}$.

Let $\mathbb{N}\{R, L\}^*$ denote the free additive (commutative) semigroup over the set $\{R, L\}^*$. The multiplication in the monoid $\{R, L\}^*$ extends by distributivity to $\mathbb{N}\{R, L\}^*$, making it the free semiring (with commutative addition and a multiplicative identity) over $\{R, L\}$. The elements of $\mathbb{N}\{R, L\}^*$ may be considered as polynomials in non-commuting indeterminates R and L with natural number coefficients. Let $A\mathbb{N}\{R, L\}^*$ denote the free right $\mathbb{N}\{R, L\}^*$-semimodule over the alphabet A. The elements of $A\mathbb{N}\{R, L\}^*$ are called *semiterms* over A [3, p. 13]. A groupoid operation $\circ$ may be defined on the set of semiterms by

$$x \circ y = xR + yL. \tag{2.3}$$

Under this groupoid operation, the subset A of $A\mathbb{N}\{R, L\}^*$ generates a subgroupoid, which is (isomorphic with) A° [3, 1.3.1]. An immediate problem is to recognize which semiterms actually lie in A°, and so represent groupoid words in A. A semiterm over A may be written in the form

$$\sum_{i=1}^{r} a_i e_i \tag{2.4}$$

with (not necessarily distinct) letters a_i and with each e_i in $\{R, L\}^*$. Ježek and Kepka then give the following criterion for recognizing groupoid words amongst semiterms.

PROPOSITION 2.1 [3, 1.3.2]. *The semiterm* (2.4) *is a groupoid word in A iff it satisfies the following three conditions:*

(i) $r \geqslant 1$;

(ii) $\forall 1 \leqslant i, j \leqslant r,\ (\exists f \in \{R, L\}^* .\ e_i = fe_j) \Rightarrow (i = j)$;

(iii) $\forall 1 \leqslant i \leqslant r,\ (\exists f, g \in \{R, L\}^* .\ \exists p \in \{R, L\} .\ e_i = fpg) \Rightarrow$
$(\exists 1 \leqslant j \leqslant r .\ \exists q \in \{R, L\} - \{p\} .\ \exists h \in \{R, L\}^* .\ e_j = hqg)$. □

The complicated-looking syntactical conditions of Proposition 2.1 may be expressed more simply in graph-theoretical language. In the leafless binary rooted tree, label the two edges growing from each vertex with R and L respectively. Each vertex of the tree is specified by the concatenated sequence of edge labels along the unique directed geodesic from the vertex to the root. A semiterm (2.4) determines a multiset $\{e_i \mid 1 \leqslant i \leqslant r\}$ of vertices in the tree. Proposition 2.1 states that the semiterm is a groupoid word w iff the multiset is the set of leaves of a rooted binary tree, a subrooted tree of the edge-labelled leafless rooted binary tree. If these conditions prevail, then the tree, with each leaf e_i labelled by a_i, is the parsing tree T_w of the groupoid word w.

Proposition 2.1 may also be restated very cleanly in code-theoretical language; namely, the semiterm (2.4) is a groupoid word w iff the multiset $\{e_i \mid 1 \leqslant i \leqslant r\}$ is a finite non-empty maximal suffix code in the alphabet $\{R, L\}$. Condition (i) of Proposition 2.1 gives the non-emptiness, condition (ii) gives the suffix property, and condition (iii) gives the maximality. If these conditions prevail, then the parsing tree T_w, shorn of its leaf labels, is the *literal representation* [1, p.87] of the code $\{e_i \mid 1 \leqslant i \leqslant r\}$. The groupoid operation (2.3) on A° corresponds to a familiar operation ([1, II(4.2)] or (2.5) below) on maximal suffix codes. In general, parsing trees of groupoid words in an alphabet A carry more information than the literal representations of maximal suffix codes. The extra information is contained in the leaf labelling by letters from A. This information vanishes if A is a singleton. In this case, groupoid words correspond exactly to finite maximal suffix codes. To summarize:

PROPOSITION 2.2. *Under the operation*

$$(C, D) \mapsto CR \cup DL, \tag{2.5}$$

finite maximal suffix codes over a binary alphabet $\{R, L\}$ form the free groupoid on one generator. □

3. Entropic Groupoids and Commutatively Prefix Codes

Over the semiring $\mathbb{N}\{R, L\}^*$ of polynomials in non-commuting indeterminates R and L with natural number coefficients, the free semimodule $A\mathbb{N}\{R, L\}^*$ on an alphabet A forms a groupoid under (2.3) in which the subset A generates the free groupoid A°. The elements of A° correspond to leaf-labelled rooted binary trees and determine

maximal suffix codes, as seen in the previous section. This section is concerned with the analogous structures involved when the indeterminates R and L commute.

Let $\mathbb{N}[R, L]$ denote the free additive (commutative) semigroup over the underlying set of the free commutative monoid over the set $\{R, L\}$. The multiplication in the monoid extends by distributivity to $\mathbb{N}[R, L]$, making it the free commutative semiring over $\{R, L\}$. The semiring $\mathbb{N}[R, L]$ may equally be taken as the semiring of polynomials in R and L with natural number coefficients. Let $A\mathbb{N}[R, L]$ denote the free $\mathbb{N}[R, L]$-semimodule over the alphabet A. Equation (2.3) may be used to define a groupoid operation on $A\mathbb{N}[R, L]$, under which the subset A generates a subgroupoid A^c.

A groupoid $(G, \circ)$ is said to be *entropic* if the binary operation $\circ: G^2 \to G$ is actually a homomorphism $\circ: (G^2, \circ) \to (G, \circ)$, i.e. if the identity

$$xy \circ zt = xz \circ yt \tag{3.1}$$

is satisfied. The groupoid $(G, \circ)$ is said to be *cancellative* if the implication

$$(xy = xz \quad \text{or} \quad yx = zx) \Rightarrow y = z \tag{3.2}$$

holds. The groupoids $(A\mathbb{N}[R, L], \circ)$ and A^c are cancellative and entropic. Moreover, A^c is the free cancellative entropic groupoid on A [3, 2.3.3], in the sense that any set map $A \to G$ into the underlying set of a cancellative entropic groupoid $(G, \circ)$ extends to a unique homomorphism $A^c \to G$.

REMARK 3.1. Although A^c is the free cancellative entropic groupoid A, it is not the free entropic groupoid on A. The identity $(a \circ xb)(yc \circ d) = (a \circ yb)(xc \circ d)$ holds in A^c, but fails with $a = b = y = d = 1$ and $x = c = 2$ in the entropic groupoid on $\{0, 1, \ldots, 6\}$ defined by $11 = 2$, $12 = 3 = 21$, $13 = 4$, $31 = 5$, $45 = 6$, and all other products 0 [2, Rem. 5.1; 3, 2.4.1]. This fact inhibits the possibility (raised in [6, §1]) of representing general entropic groupoid words faithfully by polynomials in commuting indeterminates. In some works, such as [3], the identity (3.1) is referred to as 'mediality'. The word 'entropic' is reserved in [3] to describe quotients of cancellative entropic groupoids. Such quotients are not necessarily cancellative.

Elements of the free cancellative entropic groupoid A^c on A are called *cancellative entropic groupoid words* in the letters of the alphabet A. As in the previous section, a recognition problem arises: Which elements

$$\sum_{i=1}^{r} a_i R^{n_i} L^{m_i} \tag{3.3}$$

of $A\mathbb{N}[R, L]$ represent cancellative entropic groupoid words? Ježek and Kepka give the following solution, in terms of the *total degree* $d = \max\{n_i + m_i \mid 1 \leq i \leq r\}$ of (3.3).

THEOREM 3.2 [3, 2.3.1]. *A cancellative entropic groupoid word is represented by* (3.3) *iff the following three conditions are satisfied:*

$$\left.\begin{array}{ll} \text{(i)} & r \geq 1; \\ \text{(ii)} & \forall 0 \leq m \leq n < d,\ \displaystyle\sum_{i=1}^{r} \binom{n - n_i - m_i}{m - n_i} \leq \binom{n}{m}; \\ \text{(iii)} & \forall 0 \leq m \leq d,\ \displaystyle\sum_{i=1}^{r} \binom{d - n_i - m_i}{m - n_i} = \binom{d}{m}. \end{array}\right\} \tag{3.4}$$

□

The proof of Theorem 3.2 [3, 2.3.2–8] is long and complicated, and verification of the $\frac{1}{2}(d+1)(d+2)$ conditions (3.4) for a given instance of (3.3) is tiresome. As in the previous section, however, a simpler solution to the recognition problem is offered by coding theory. Note that (3.3) represents a cancellative entropic groupoid word iff there is a groupoid word (2.4) for which each e_i maps to $R^{n_i}L^{m_i}$ under the homomorphism k, from the free monoid $\{R, L\}^*$ on $\{R, L\}$ to the free commutative monoid on $\{R, L\}$, induced by the identity set map on $\{R, L\}$. In terms of coding theory, this means that the multiset $\{R^{n_i}L^{m_i} \mid 1 \leqslant i \leqslant r\}$ is commutatively equivalent to a multiset $\{e_i \mid 1 \leqslant i \leqslant r\}$ that is a (finite) maximal suffix code; in other words, is maximal and commutatively suffix or maximal and commutatively prefix [1, §VII.6]. Using a well-known criterion for such multisets [4, Th. 1; 1, Th. VIII.6.1], one obtains the following solution to the recognition problem as an alternative to Theorem 3.2.

THEOREM 3.3. *A cancellative entropic groupoid word is represented by* (3.3) *iff the following condition is satisfied:*

$$\exists Q(R, L) \in \mathbb{N}[R, L] \,.\, \sum_{i=1}^{r} R^{n_i}L^{m_i} = 1 + (R + L - 1)Q(R, L). \tag{3.5}$$

□

COROLLARY 3.4. *The conditions* (3.4) *and* (3.5) *on* (3.3) *or the polynomial* $\sum_{i=1}^{r} R^{n_i}L^{m_i}$ *are equivalent.* □

4. STRUCTURE FUNCTIONS AND THE THERMODYNAMIC INTERPRETATION

The polynomial

$$P(R, L) = \sum_{i=1}^{r} R^{n_i}L^{m_i} \tag{4.1}$$

appearing on the left-hand side of the equation in (3.5) is called the *bivariate structure function* of the element (3.3) of $A\mathbb{N}[R, L]$. Similarly, one may associate a bivariate structure function

$$P(R, L) = \sum_{i=1}^{r} e_i k \tag{4.2}$$

with a semiterm (2.4), where k is the homomorphism from the free monoid to the free commutative monoid defined in the previous section. ([4, §II] associates 'multivariate structure functions' with subsets of free monoids over arbitrary finite alphabets.) In addition to the criterion [4, Th. 1] in terms of bivariate structure functions that is embodied in Theorem 3.3, Karp gave a criterion in terms of functions of a single variable, the 'univariate structure functions' [4, Th. 2]. These functions were defined in terms of 'costs'. However, when reinterpreted in terms of 'energies', with the single variable reformulated in terms of 'temperature', the univariate structure function turns out to be a basic and familiar function from thermodynamics, the 'Zustandsumme', 'sum-over-states' or *partition function* [9, pp. 532, 567] of Darwin and Planck.

As a substrate for the physical interpretation, consider the first quadrant $\mathbb{N}^2$ of the lattice $\mathbb{Z}^2$. A particle may start at the origin $(0, 0)$ with zero energy. Each move from (n, m) to $(n+1, m)$ increases its energy by a quantity ρ. Each move from (n, m) to $(n, m+1)$ increases its energy by the quantity λ. Thus a particle at position (n, m) has

energy $n\rho + m\lambda$. The element (3.3) of $A\mathbb{N}[R, L]$ represents a system with r states, the ith state corresponding to a particle at position (n_i, m_i), and thus having the energy $n_i\rho + m_i\lambda$ assigned to it. The partition function associated with the system (3.3) is then

$$Z(T) = \sum_{i=1}^{r} \exp\{-(n_i\rho + m_i\lambda)/kT\}, \tag{4.3}$$

where T is the temperature and k is Boltzmann's constant. Karp's univariate structure function $F(z)$ is $Z(-1/k \log z)$, restricted to the case of positive integral 'energies' or 'costs' ρ, λ so that $F(z)$ becomes an element of $\mathbb{N}[z]$, a polynomial in z with natural number coefficients. Using the criterion [4, Th. 2] on univariate structure functions, the solution to the recognition problem for cancellative entropic groupoid words may be formulated as follows.

THEOREM 4.1. *Let the energies ρ, λ be positive integers. Then the element* (3.3) *of* $A\mathbb{N}[R, L]$ *represents a cancellative entropic groupoid word iff its partition function* (4.3) *is of the form*

$$1 + (e^{-\rho/kT} + e^{-\lambda/kT} - 1)Y(T),$$

for some element $Y(T)$ *of* $\mathbb{N}[e^{-1/kT}]$. □

5. Convexity and Binary Modes

A groupoid $(G, \circ)$ is said to be *idempotent* if each singleton in G is a subgroupoid, i.e. if the identity

$$x \circ x = x \tag{5.1}$$

is satisfied. Algebras (of arbitrary type) that are both idempotent and entropic, i.e. having all singletons as subalgebras and all operations homomorphic, are referred to as *modes* [7, pp. vi, 14]. In general, the theory of modes is closely connected with the concept of convexity. Idempotent, entropic groupoids, with their single binary operation, are *binary modes.*

Let $\mathbb{Z}[M]$ be the integral polynomial ring in the indeterminate M. Given an alphabet A, let $A\mathbb{Z}[M]$ be the free A-module over $\mathbb{Z}[M]$. Under the binary operation

$$x \circ y = xM + y(1 - M), \tag{5.2}$$

$A\mathbb{Z}[M]$ becomes a binary mode. Let A^b denote the submode generated by A. Then A^b is the free binary mode on the set A (this follows from [2], or from [3, 2.4.4] and the freeness of A^c as a cancellative entropic groupoid). The elements of A^b are described as *binary mode words* in the alphabet A. An arbitrary element of $A\mathbb{Z}[M]$ may be written in the form

$$\sum_{i=1}^{s} b_i p_i(M) \tag{5.3}$$

with $\{b_1, \ldots, b_s\}$ as an s-element subset of A and $\{p_i(M) \mid 1 \leqslant i \leqslant s\}$ as a subset of $\mathbb{Z}[M]$. As before, a recognition problem arises: When does (5.3) represent a binary mode word in A? Solutions were given by van Maaren in his Utrecht dissertation [5] under Monna's direction. The most immediate solution is a combinatorial one, analogous to Theorem 3.2:

THEOREM 5.1. *A binary word is represented by* (5.3) *iff the following three conditions are satisfied:*

$$\left.\begin{array}{ll} \text{(i)} & s \geqslant 1; \\ \text{(ii)} & \forall 1 \leqslant i \leqslant s, \quad \exists N_i \in \mathbb{N}\,.\ \forall 1 \leqslant j \leqslant N_i, \quad \exists 0 \leqslant A_{ij} \leqslant \binom{N_i}{j} \\ & p_i(M) = \sum_{j=1}^{N_i} A_{ij} M^j (1-M)^{N_i - j}; \\ \text{(iii)} & \sum_{i=1}^{s} p_i(M) = 1. \end{array}\right\} \tag{5.4}$$

PROOF. See [5, Th. II.4.1] and [5, Prop. II.2.3]. □

An alternative solution to the recognition problem may be couched in geometric language, indicating one of the connections between modal theory and convexity. Let T denote the convex hull of the points $(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ in the Euclidean space $\mathbb{R}^s$.

THEOREM 5.2. *A binary mode word is represented by* (5.3) *iff*

$$[0, 1] \to T;\ t \mapsto (p_1(t), p_2(t), \ldots, p_s(t)) \tag{5.5}$$

parametrizes a directed curve in the geometric simplex T leading from one extreme point to another.

PROOF. See [5, Th. II.4.11] and [5, Th. II.4.9]. □

6. FINITE MAXIMAL CODES

Let $m: \mathbb{Z}[R, L] \to \mathbb{Z}[M]$ denote the ring morphism of polynomial rings induced by $R \mapsto M$ and $L \mapsto (1 - M)$. For a real number q, let $\eta_q: \mathbb{Z}[M] \to \mathbb{R}$ denote the ring morphism induced by $M \mapsto q$. For q in the unit interval, the composite

$$\{R, L\}^* \xrightarrow{k} \mathbb{Z}[R, L] \xrightarrow{m} \mathbb{Z}[M] \xrightarrow{\eta_q} \mathbb{R} \tag{6.1}$$

is the *Bernoulli distribution* π on $\{R, L\}^*$ with $\pi(R) = q$ and $\pi(L) = 1 - q$ [1, p. 54]. For an s-element subset ('language')

$$X = \{e_i \mid 1 \leqslant i \leqslant s\} \tag{6.2}$$

of $\{R, L\}^*$, one has

$$\pi(X) = \sum_{i=1}^{s} e_i k m \eta_q \tag{6.3}$$

as the *measure of X relative to* π [1, p. 55]. Since attention focuses on languages of measure 1, the following equivalence is worth noting.

PROPOSITION 6.1 *For a finite language X as in* (6.2), *the following are equivalent:*
(i) $\pi(X) = 1$ *for all Bernoulli distributions* π;
(ii) $\pi(X) = 1$ *for a Bernoulli distribution assigning transcendental weight q to the letter R*;
(iii) $\sum_{i=1}^{s} e_i k m = 1$ *in* $\mathbb{Z}[M]$.

PROOF. (i)$\Rightarrow$(ii) is trivial.

(ii)$\Rightarrow$(iii): Since q is transcendental, the ring homomorphism $\eta_q\colon \mathbb{Z}[M]\to\mathbb{R}$ injects. Then $1\eta_q = 1 = \pi(X) = \sum_{i=1}^{s} e_i km\eta_q$ implies $1 = \sum_{i=1}^{s} e_i km$.

(iii)$\Rightarrow$(i): If π assigns weight q to R, then $\sum_{i=1}^{s} e_i km = 1$ implies $\pi(X) = \sum_{i=1}^{s} e_i km\eta_q = 1\eta_q = 1$. □

Any groupoid word may be construed as a cancellative entropic groupoid word, and any cancellative entropic groupoid word may be construed as a binary mode word. These constructions are reflected by extensions of the morphisms k and m for a given alphabet A, namely

$$k\colon A\mathbb{N}\{R, L\}^*\to A\mathbb{N}[R, L];\ \sum_{i=1}^{r} a_i e_i \mapsto \sum_{i=1}^{r} a_i e_i k \tag{6.4}$$

and

$$m\colon A\mathbb{N}[R, L]\to A\mathbb{Z}[M];\ \sum_{i=1}^{r} a_i R^{n_i} L^{m_i} \mapsto \sum_{i=1}^{r} a_i M^{n_i}(1-M)^{m_i}, \tag{6.5}$$

together with their groupoid homomorphism restrictions

$$k\colon A^\circ\to A^c \quad \text{and} \quad m\colon A^c\to A^b. \tag{6.6, 6.7}$$

Conversely, any cancellative entropic groupoid word may be construed as one or more groupoid words, e.g. $ab\circ cd$ may be construed as $ab\circ cd$ or $ac\circ bd$. Similarly, any binary mode word may be construed as one or more cancellative entropic groupoid words, e.g. $a\circ bc$ may be construed as $aa\circ bc$ or $ab\circ ac$ or $a\circ bc$. However, it should be noted that the recognition criteria discussed above cannot be lifted back along the maps k of (6.4) and m of (6.5) *in general*:

EXAMPLE 6.2. (i) Consider the semiterm $aR + bLR + cL^2$, which is not a groupoid word since $\{R, LR, L^2\}$ is not a suffix code. The image of the semiterm under (6.4) is the cancellative entropic groupoid word $(aR + bLR + cL^2)k = aR + bRL + cL^2 = a\circ(b\circ c)$. Of course, $\{R, LR, L^2\}$ is commutatively suffix without being suffix.

(ii) (cf. [4, Ex. (c) to Th. 1]) Consider the element $a(R^3 + L^3 + 3RL)$ of $A\mathbb{N}[R, L]$. It is not a cancellative entropic groupoid word, since $R^3 + L^3 + 3RL = 1 + (R + L - 1)(R^2 + L^2 - RL + R + L + 1)$. Nevertheless, its image under (6.5) is the (short) binary mode word a. A more illuminating version of the example is offered by the element $aR^3 + bL^3 + cRL + dRL + eRL$ of $A\mathbb{N}[R, L]$. Again, this is not a cancellative entropic groupoid word, although it maps under (6.5) to the binary mode word $(ac\circ d)(e\circ cb)$.

Suppose that the language X of (6.2) is a finite maximal code. By [1, Th. 5.10], X satisfies condition (ii) of Proposition 6.1. Let $B = \{b_i \mid 1\leqslant i\leqslant s\}$ be an s-element alphabet. Since each $e_i km$ is a product of powers of M and $(1 - M)$, and $\sum_{i=1}^{s} e_i km = 1$ by Proposition 6.1, Theorem 5.2 shows that the semiterm

$$\sum_{i=1}^{s} b_i e_i km \tag{6.8}$$

is a binary mode word in B. In this sense, finite maximal codes determine binary mode words. If the recognition criteria for groupoid words could be lifted back along m *in such a case,* so that $\sum_{i=1}^{s} b_i e_i k$ represented a cancellative entropic groupoid word, it would then follow that the finite maximal code X was commutatively prefix. Thus the

correspondences between codes and groupoid words discussed in this paper offer an approach to 'the main question left open in the theory of codes' [1, p. 423], the question as to whether every finite maximal code is commutatively prefix.

References

1. J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, Orlando, Florida, 1985.
2. S. Fajtlowicz and J. Mycielski, On convex linear forms, *Alg. Univ.*, **4** (1974), 244–249.
3. J. Ježek and T. Kepka, *Medial Groupoids*, Academia, Praha, 1983.
4. R. M. Karp, Minimum-redundancy coding for the discrete noiseless channel, *IRE Trans. Inform. Theory*, **IT-7** (1961), 27–38.
5. H. van Maaren, Algebraic simplices in modules, Doctoral thesis, Rijksuniversiteit te Utrecht, January 1979.
6. H. Minc, Index polynomials and bifurcating root-trees, *Proc. R. Soc. Edinb.*, **A64** (1957), 319–341.
7. A. B. Romanowska and J. D. H. Smith, *Modal theory*, Heldermann, Berlin, 1985.
8. J. D. H. Smith, Skein polynomials and entropic right quasigroups, *Demonstr. Math.* to appear.
9. R. C. Tolman, *The Principles of Statistical Mechanics*, Dover, New York, 1979.
10. E. Witten, Quantum field theory and the Jones polynomial, *Commun. Math. Phys.*, **121** (1989), 251–399.

Received 10 September 1991 and accepted in revised form 12 January 1991

Jonathan D. H. Smith
Department of Mathematics, 400 Carver,
Iowa State University,
Ames, Iowa 50011, U.S.A.